# Visualizing "Potential Student" Data Using A
# Survey Microdata

SHEEO Communities of Practice Workshop

Washington, D
September 27, 20

Nate Johnson

Postsecondary Analy
423 East Virginia St
Tallahassee, Florida 3
8502940672

www.postsecondaryanalyti
nate.johnson@postsecondaryana

Postsecondary Analytics

# Objectives

By end of workshop, participants should …

- Know why American Community Survey microdata is useful to their agencies' mission
- Be excited to start using the data in their state
- Be confident that they can get and use the data on their own
- Understand the major limitations of the data

# Why ACS Microdata is Important

- States usually have good information about students
- But how helpful is that for understanding who isn't being well-served?
- Attainment goals are about reaching non-students (potential students or ex-students)
- ACS is a good source of data for underserved populations who are not reflected in student data systems

# What's in ACS data?

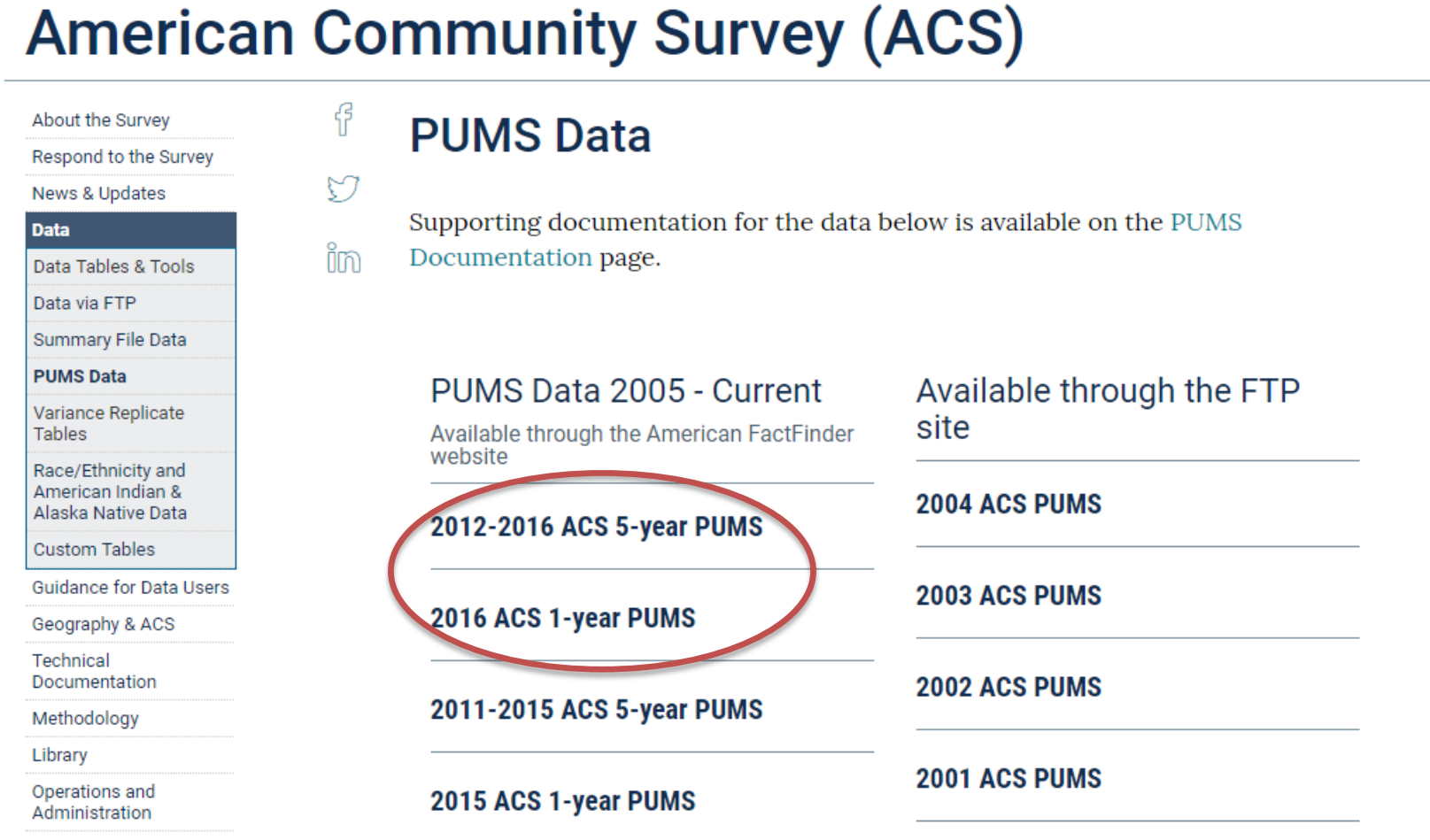
- Basic variables for attainment:
  - Highest attainment level
  - Enrollment status (K-12, postsecondary, public/private)
  - Field of degree (bachelor's only)
- Demographic characteristics
  - Age (down to single year)
  - Race / ethnicity (detailed)
  - Gender
  - Language
  - Citizenship
  - Disability status
  - Family composition
- Employment & income
  - Status
  - Hours, wages, weeks
  - Occupation
  - Income sources
  - Federal benefits
  - Military/veteran status

# What's in ACS data?

- Geographic
  - Public Use Microdata Area (PUMA) = ~100,000 population
  - Migration within 1 year (same or other state/PUMA)
  - Commute distance & time
- Household
  - Composition
  - Housing type (detailed)
  - Housing finance (value, costs, taxes)
  - Ownership status
  - Broadband/computer access
  - Household demographics
  - Household income

# Where to Get the Data

https://www.census.gov/programs-surveys/acs/data/pums.html



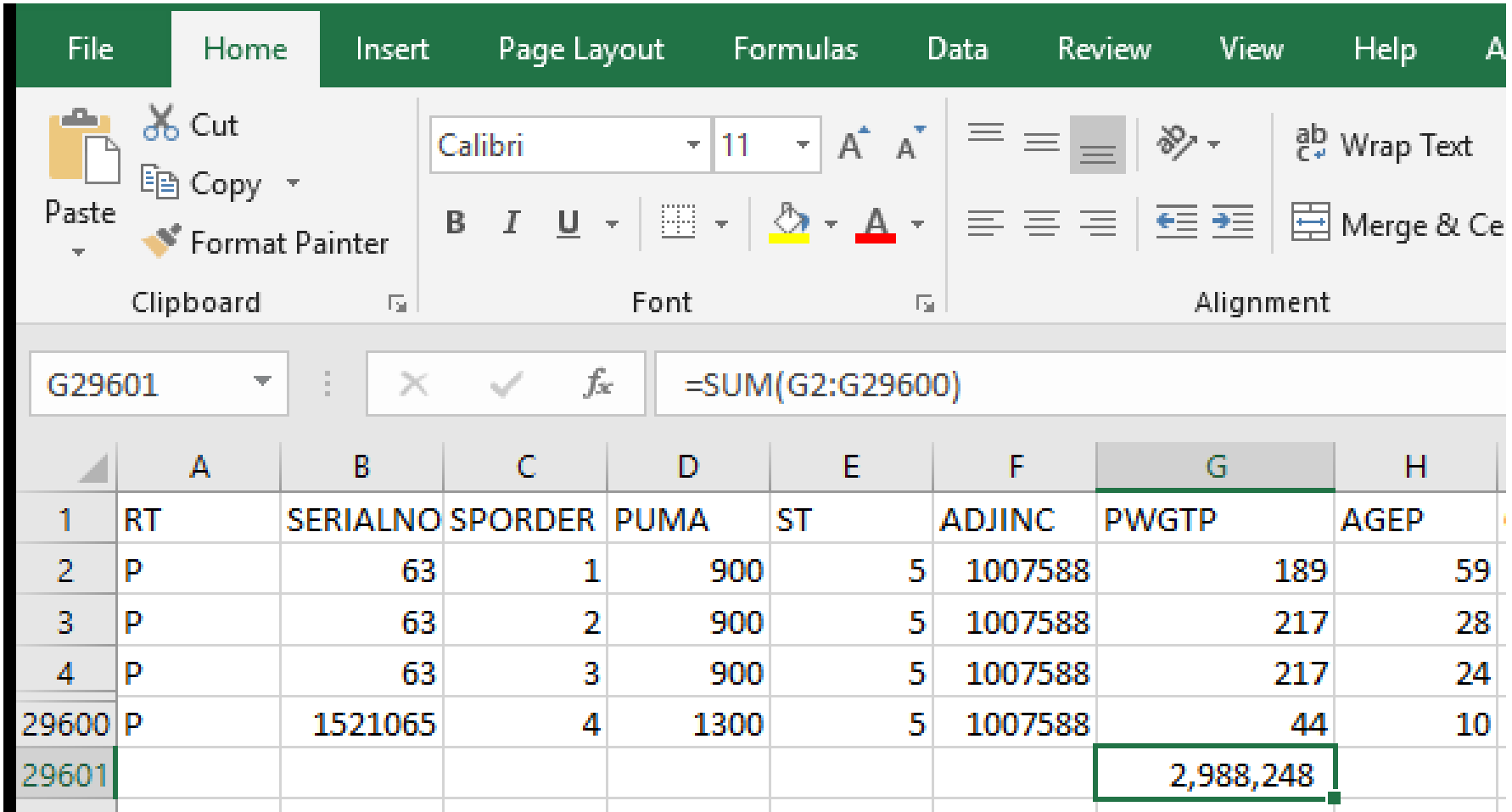
# Where to Get the Data

https://www.census.gov/programs-surveys/acs/data/pums.html

| United States Population Records | United States Housing Unit Records |
|---|---|
| Alabama Population Records | Alabama Housing Unit Records |
| Alaska Population Records | Alaska Housing Unit Records |
| Arizona Population Records | Arizona Housing Unit Records |
| Arkansas Population Records | Arkansas Housing Unit Records |
| California Population Records | California Housing Unit Records |
| Colorado Population Records | Colorado Housing Unit Records |
| Connecticut Population Records | Connecticut Housing Unit Records |
| Delaware Population Records | Delaware Housing Unit Records |
| District of Columbia Population Records | District of Columbia Housing Unit Records |
| Florida Population Records | Florida Housing Unit Records |
| Georgia Population Records | Georgia Housing Unit Records |

# Diving into the Data: Arkansas Example

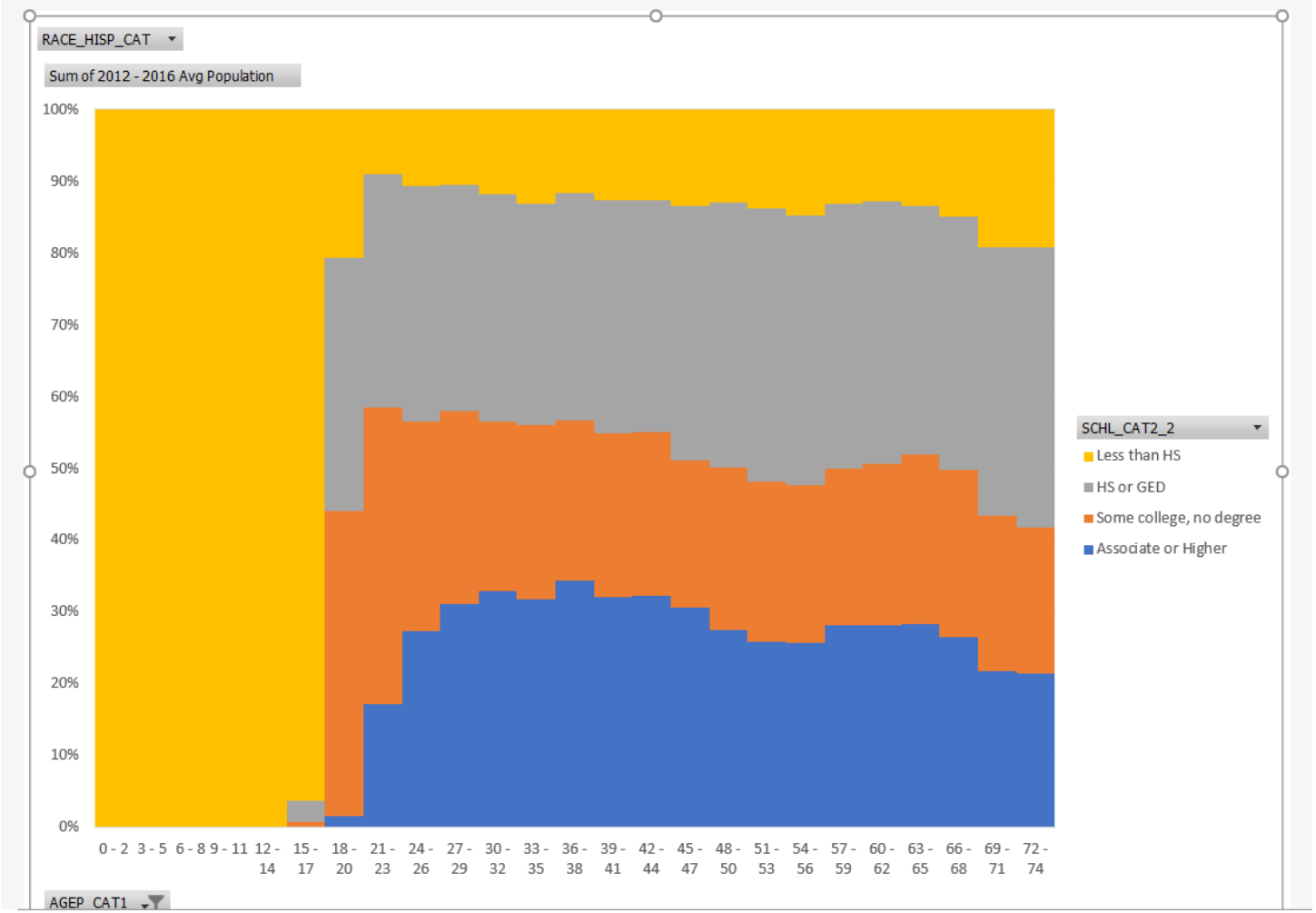
- Add up the population weights for the groups of interest
- Total of "PWGTP" = Total 2016 State Population




| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | RT | SERIALNO | SPORDER | PUMA | ST | ADJINC | PWGTP | AGEP |
| 2 | P | 63 | 1 | 900 | 5 | 1007588 | 189 | 59 |
| 3 | P | 63 | 2 | 900 | 5 | 1007588 | 217 | 28 |
| 4 | P | 63 | 3 | 900 | 5 | 1007588 | 217 | 24 |
| 29600 | P | 1521065 | 4 | 1300 | 5 | 1007588 | 44 | 10 |
| 29601 | | | | | | | 2,988,248 | |

# Arkansas Example

- Using ACS Microdata in Excel {see workbook for Arkansas example}



# Arkansas Example – Excel Demonstration

- Using the population variable
- Identifying the target / potential student population
- Attainment by age
- Detailed race/ethnicity data

# “Lightly Processed” State Files in Drop

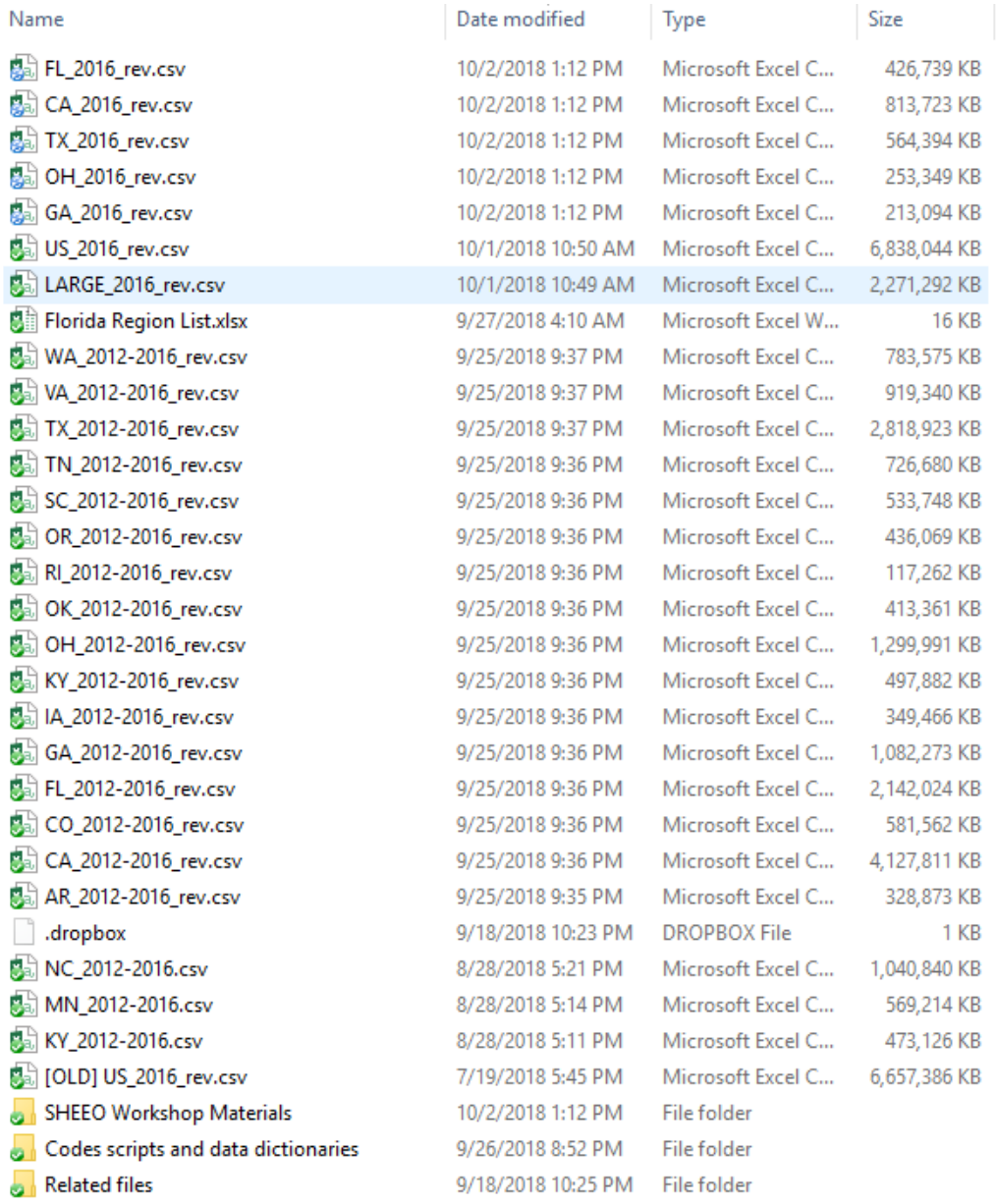
- Available at:

https://www.dropbox.com/sh/hjjjyahwep2f322/AAARLMBhjGhUYJgyToVQwjXIa?dl=0

- Single files for easier use in Excel or Tableau
- Selects a subset of the most relevant variables
- Labels some values and creates some important groupings (e.g. combined Race/Ethnicity)
- Links key variables from “Household” record (e.g. household income) to the “Person” record (individual file)
- Currently has
  - 2012-2016 five-year files for participating states
  - Single year 2016 file for all 50 states in one file
  - Single year 2016 file for large states: CA, TX, FL, OH, GA
- SAS code to create files also in Dropbox

# “Lightly Processed” State Files in Drop

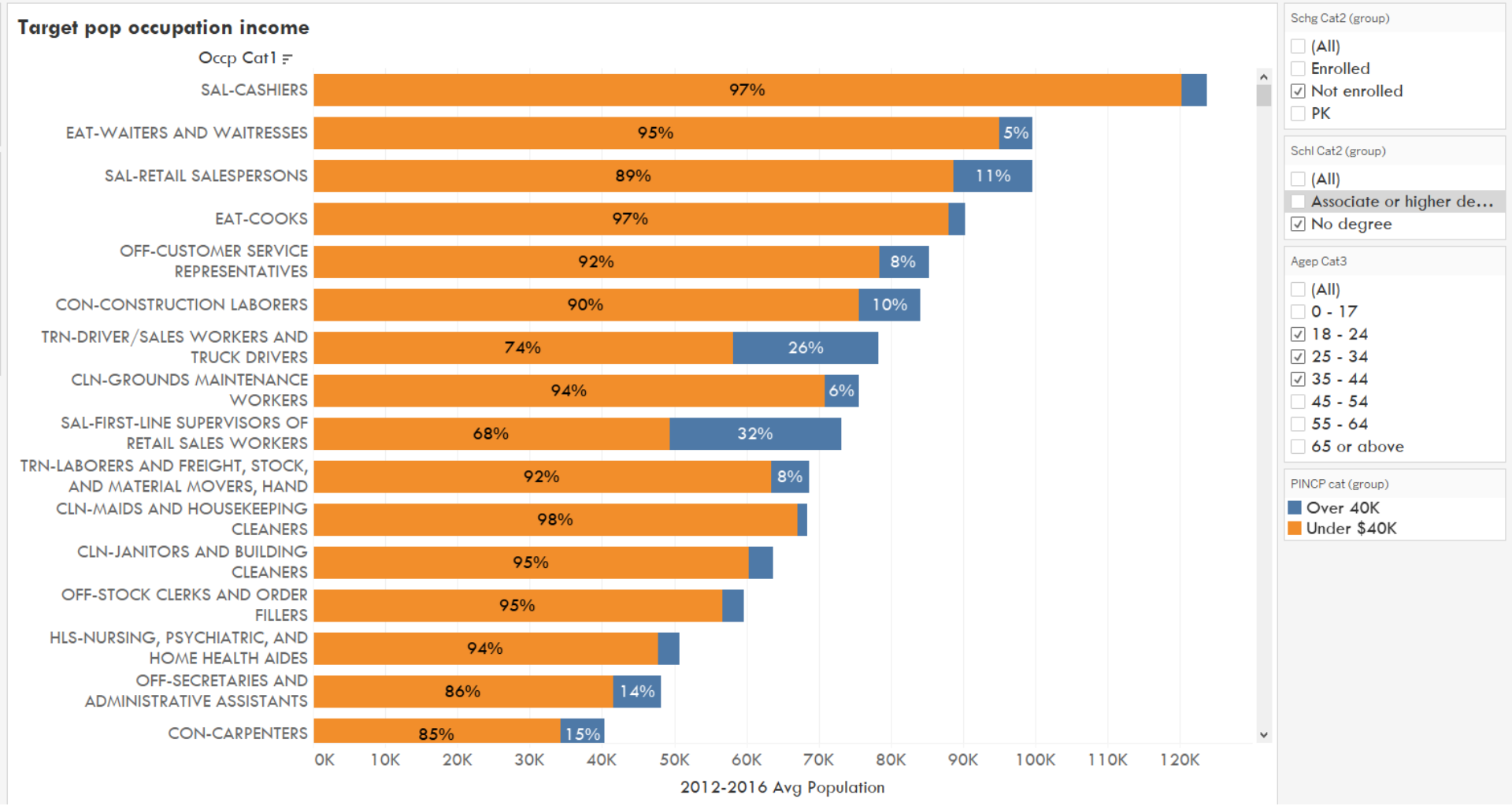
- Available files in Dropbox as of Oct 1, 2018
- Available at least through Nov. 1, 2018
- If you use these, please download to your own computer first



| Name | Date modified | Type | Size |
|---|---|---|---|
| FL_2016_rev.csv | 10/2/2018 1:12 PM | Microsoft Excel C... | 426,739 KB |
| CA_2016_rev.csv | 10/2/2018 1:12 PM | Microsoft Excel C... | 813,723 KB |
| TX_2016_rev.csv | 10/2/2018 1:12 PM | Microsoft Excel C... | 564,394 KB |
| OH_2016_rev.csv | 10/2/2018 1:12 PM | Microsoft Excel C... | 253,349 KB |
| GA_2016_rev.csv | 10/2/2018 1:12 PM | Microsoft Excel C... | 213,094 KB |
| US_2016_rev.csv | 10/1/2018 10:50 AM | Microsoft Excel C... | 6,838,044 KB |
| LARGE_2016_rev.csv | 10/1/2018 10:49 AM | Microsoft Excel C... | 2,271,292 KB |
| Florida Region List.xlsx | 9/27/2018 4:10 AM | Microsoft Excel W... | 16 KB |
| WA_2012-2016_rev.csv | 9/25/2018 9:37 PM | Microsoft Excel C... | 783,575 KB |
| VA_2012-2016_rev.csv | 9/25/2018 9:37 PM | Microsoft Excel C... | 919,340 KB |
| TX_2012-2016_rev.csv | 9/25/2018 9:37 PM | Microsoft Excel C... | 2,818,923 KB |
| TN_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 726,680 KB |
| SC_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 533,748 KB |
| OR_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 436,069 KB |
| RI_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 117,262 KB |
| OK_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 413,361 KB |
| OH_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 1,299,991 KB |
| KY_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 497,882 KB |
| IA_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 349,466 KB |
| GA_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 1,082,273 KB |
| FL_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 2,142,024 KB |
| CO_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 581,562 KB |
| CA_2012-2016_rev.csv | 9/25/2018 9:36 PM | Microsoft Excel C... | 4,127,811 KB |
| AR_2012-2016_rev.csv | 9/25/2018 9:35 PM | Microsoft Excel C... | 328,873 KB |
| .dropbox | 9/18/2018 10:23 PM | DROPBOX File | 1 KB |
| NC_2012-2016.csv | 8/28/2018 5:21 PM | Microsoft Excel C... | 1,040,840 KB |
| MN_2012-2016.csv | 8/28/2018 5:14 PM | Microsoft Excel C... | 569,214 KB |
| KY_2012-2016.csv | 8/28/2018 5:11 PM | Microsoft Excel C... | 473,126 KB |
| [OLD] US_2016_rev.csv | 7/19/2018 5:45 PM | Microsoft Excel C... | 6,657,386 KB |
| SHEEO Workshop Materials | 10/2/2018 1:12 PM | File folder | |
| Codes scripts and data dictionaries | 9/26/2018 8:52 PM | File folder | |
| Related files | 9/18/2018 10:25 PM | File folder | |

# Florida Example

- Using ACS Microdata in Tableau {see workbook for Florida example}



# Florida Example – Tableau Demonstration

- Attainment by race / ethnicity with additional detail
- "Potential student" target population
  - Employment & income
  - Children
  - Aid eligibility
  - Occupation
  - Federal benefits

# 2016 or 2012-16?

- 2016 (1% sample) is best for
  - Larger states
  - Most recent data available
  - When data are not highly disaggregated
- 2012-16 (5% sample) is best for
  - Smaller states
  - Averages across longer time horizons (2020 may be more like 2014 than like 2016)
  - More disaggregated analyses
  - Divide population weight by 5

# Sampling & Error

- Rules of thumb: don't disaggregate below:
  - 10,000 in 1-year file
  - 2,000 in 5-year file

EXAMPLE

- # of Black 19-Year-Olds in Arkansas in 2016
  - Estimate: 10,011 (21% of all 19-year olds)
  - 90% confidence interval: 7,980-12,042 (17%-25%)
  - OK as rough estimate.
- # of Black 19-Year-Olds in Arkansas with 1+ years of college and no degree
  - Estimate: 3,828 (38% of all black 19-year-olds)
  - 90% confidence interval: 2,232-5,424 (22%-54%)
  - Not especially helpful.
- Want more? Love formulas? See "Readme" file in Dropbox

# Additional Topics

What else would be helpful?

- Geography/mapping
- Confidence intervals/standard error estimates
- Income variable adjustments
- Population estimates/projections/adjustments
- Excel or Tableau features
- Weighted averages
- Other?

# Drilling Down by Geography

## 2010 Census Public Use Microdata Area (PUMA) Reference Maps - Florida

The total number of map sheets is listed next to each entity name. In instances where there is only one map sheet for a given entity, the map link will directory that contains all of the maps sheets for that entity.

| Code | Name | Total Map Sheets |
|---|---|---|
| 12 00101 | Alachua County (Central)--Gainesville City (Central) PUMA | 1 |
| 12 00102 | Alachua County (Outer) PUMA | 1 |
| 12 00500 | Walton, Washington, Holmes & Bay Counties PUMA | 1 |
| 12 00901 | Brevard County (Northwest)--Titusville, Rockledge & Cocoa Cities PUMA | 1 |
| 12 00902 | Brevard County (East)--Beaches & Merritt Island PUMA | 1 |
| 12 00903 | Brevard County (Southwest)--Melbourne & West Melbourne Cities PUMA | 1 |
| 12 00904 | Brevard County (Southeast)--Palm Bay City, Grant-Valkaria & Malabar Towns PUMA | 1 |
| 12 01101 | Broward County (West)--Coral Springs & Parkland Cities PUMA | 1 |
| 12 01102 | Broward County (North Central)--Margate & Coconut Creek Cities PUMA | 1 |
| 12 01103 | Broward County--Deerfield, Pompano Beach (North) & Lighthouse Point Cities PUMA | 1 |
| 12 01104 | Broward County--Pompano Beach (South) & Fort Lauderdale (Northeast) Cities PUMA | 1 |
| 12 01105 | Broward County (Central)--Tamarac, Oakland Park & North Lauderdale Cities PUMA | 1 |
| 12 01106 | Broward County (Central)--Plantation & Sunrise Cities PUMA | 1 |
| 12 01107 | Broward County (Central)--Lauderhill & Lauderdale Lakes Cities PUMA | 1 |
| 12 01108 | Broward County (East Central)--Fort Lauderdale City (Central) PUMA | 1 |
| 12 01109 | Broward County (Southeast)--Hollywood (North) & Dania Beach (South) Cities PUMA | 1 |
| 12 01110 | Broward County (Central)--Davie Town & Cooper City PUMA | 1 |
| 12 01111 | Broward County--Weston, Pembroke Pines (Northwest) Cities & Southwest Ranches Town PUMA | 1 |
| 12 01112 | Broward County (South Central)--Miramar (West) & Pembroke Pines (Southwest) Cities PUMA | 1 |

# Formula for Replicate Estimate Standa

The standard error of X can be approximated after the replicate estimates $X_1$ through $X_{80}$ are computed. The standard error is estimated using the sum of squared differences between each replicate estimate $X_r$ and the full sample estimate X. The standard error formula is:

$$SE(X) = \sqrt{\frac{4}{80}\sum_{r=1}^{80}(X_r - X)^2}$$

If X is zero, then use the generalized variance method for zero estimates given in Section 6.2.a, Standard Errors for Totals and Percentages, to approximate the standard error.

From ACS documentation “README” file.